



Oracle Storage Management: New Techniques and Choices



THE DATA AVAILABILITY COMPANY™

May We Suggest...

- ▼ Cell phones and pagers on *vibrate* or *silent*, please



Data
Access

Data
Protection

Issues in Oracle Storage Management

- ▼ Oracle I/O is a, uh, rich mixture of I/O types
- ▼ Physical database layout issues causing outages
 - To your system
 - To your personal/social/family life
- ▼ S.A.M.E. methodology recommended
 - But how to implement?
- ▼ Interest in OPS tempered by egregious system vendor requirements
- ▼ Many people's backup in the stone age

Data
Access

Data
Protection

What's in Oracle *9i*

▼ Applications Server

- Available now
- Tools tools for web sites
 - Apache
 - Caching
 - Wireless access
 - Forms
 - Portals and portlets
 - Personalization
 - Clickstream analyzer/Business Intelligence
 - etc.

▼ Database

- Target: Mid-year 2001
- RAC (Real Application Clusters) replaces OPS
 - Cache fusion
 - Systems Management
- DataSafe
- Online Modification
- Schema changes
 - Index operations
 - DB changes
- Security enhancements
- etc.
- and...

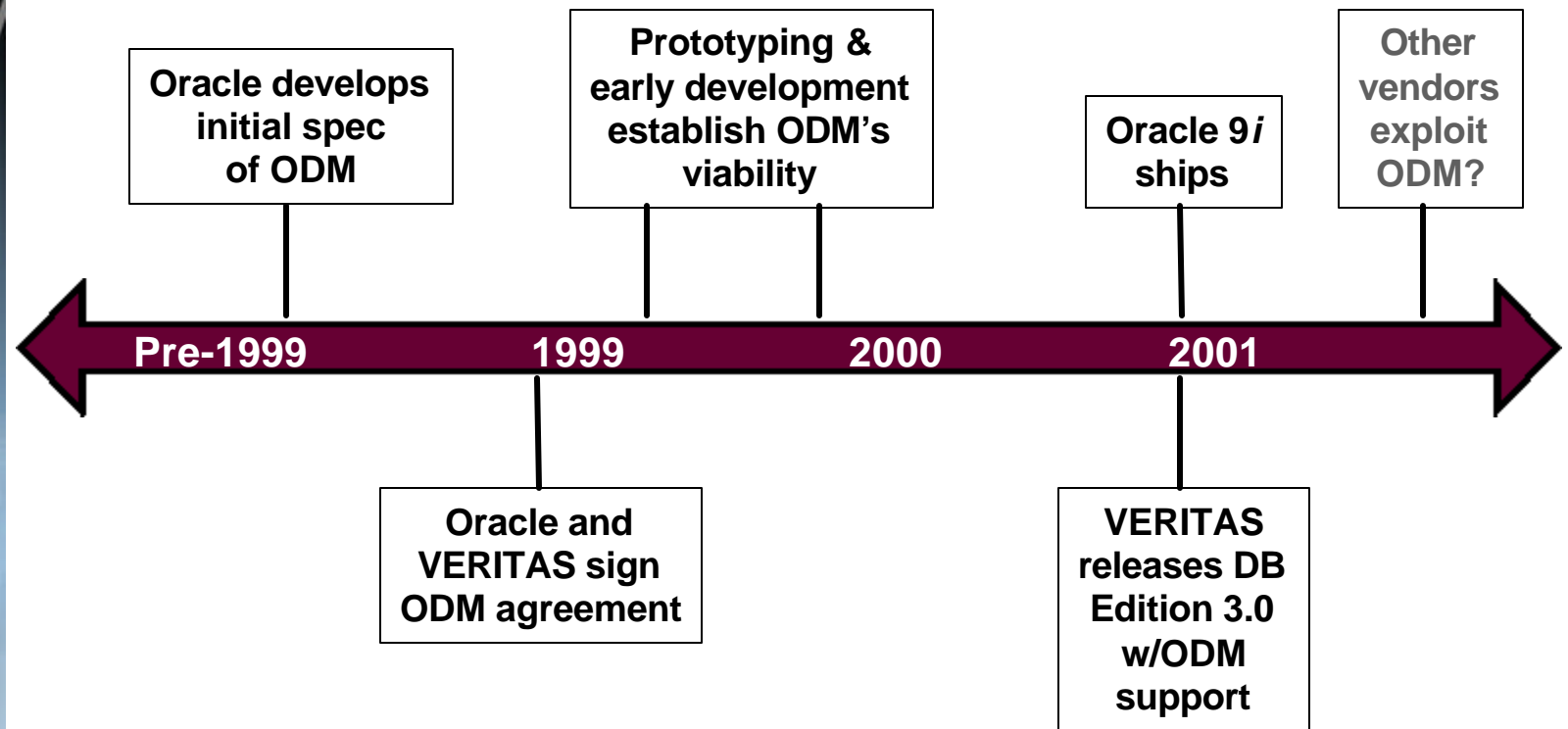
Sweet!

**Oracle Disk Manager
(ODM)**

Data
Access

Data
Protection

ODM History



Data
Access

Data
Protection

Feature Overview

- ▼ **Full Featured IO - Compliant with Oracle Corporation Requirements**
- ▼ **File Creation Features**
- ▼ **File Resizing Features**
- ▼ **File Identification Feature**
- ▼ **Special Application - S.A.M.E**

Data
Access

Data
Protection

Oracle File IO Basics

- ▼ **Oracle Server generates a very complex IO mix**
 - **scattered reads**
 - Shadow Processes, MTS.
 - Some PQO index range scans, etc
 - **single-block reads**
 - Shadow Processes, MTS, PQO.
 - **direct reads**
 - Shadow Processes. Reads from TEMP segments, NOCACHE hint, etc
 - PQO Scan Processes
 - **single-block asynchronous writes**
 - DBWR processes. Write batches bound for many files, need to be asynchronous for performance's sake
 - **direct writes**
 - Shadow Processes. Writes to TEMP segments, direct path loads, etc
 - PQO writes
 - **large asynchronous writes**
 - lgwr writes - if redo buffer is larger than the port-level max supported IO, LGWR will issue multiple asynchronous writes to flush the buffer

Data
Access

Data
Protection



ODM File I/O

▼ Oracle without ODM

- Without ODM, Oracle must resort to many different sets of calls to provide the wide variety of IO types.
 - Example:
`pwrite()/pread(), async_write(), readv(), read(), write()
() lio_listio(), kaio()`
- Without ODM, asynchronous DBWR page flushing requires two calls: one to issue the I/O and another to poll for completed IO.

▼ Oracle with ODM

- With ODM, Oracle needs only a single call
 - `odm_io()`
 - `odm_io()` supports ALL Oracle file IO types on ALL files (Raw or VxFS)!
 - Gathered writes (DBWR) and LGWR asynchronous writes occur with a single call to `odm_io()` without regard for file type (VxFS or RAW) or number of target files

Data
Access

Data
Protection

ODM File Creation Features

- ▼ ODM Includes features that enable more effective Oracle file creation
- ▼ Without ODM, failed attempts to add files to a database will result in an unused file to be cleaned up from outside Oracle.
- ▼ With ODM, files are no longer created with traditional `open()` or `creat()`
 - Files will be created with `odm_create()` and then initialized or filled.
 - If Oracle is happy with the file it calls `odm_commit()`
 - If Oracle is not happy with the file, Oracle calls `odm_abort()`. The file will be completely cleaned up from within ODM.
- ▼ With ODM, VxFS files are laid out with contiguous disk blocks.
 - This aids table and index scan throughput

Data
Access

Data
Protection

File Resizing Features

- ▼ **ODM supports Oracle's autoextend by adding contiguous disk blocks to a tablespace**
 - However, there will most likely be a hole between the old datafile and the extents added with the `odm_resize()` command
- ▼ This is not adding a datafile, but instead just adding extents to the existing datafile.
 - Oracle cannot autoextend a raw partition unless it is a VxVM volume
 - On UFS, the added portion will not be contiguous

Data
Access

Data
Protection

VERITAS
BUSINESS WITHOUT INTERRUPTION™



Data
Access



Data
Protection



File Identification Feature

- ▼ Large databases require too many kernel resources. Most notably per-process and system-wide file descriptor slots
- ▼ With Oracle Dedicated Processes (Shadow Processes), large numbers of users combined with large numbers of datafiles becomes an issue.
 - Example: 1000 users on 500 datafiles will cost 500,000 kernel file descriptor table entries
 - Note, the math is not really based on “Users”. Each Shadow process may need to open every datafile. Oracle7 shadow process open all datafiles when they connect. Oracle8 and beyond open files upon their first reference.
 - A “User” may actually have many shadow processes depending upon the application.
- ▼ Kernel file descriptor usage is protected with locking. Performing a high rate of implicit/explicit file open/close operations will cause undesirable amount of kernel-mode CPU cycle wastage.
 - File open/close internals use the same kernel locking infrastructure.

Data
Access

Data
Protection

File Identification Feature (cont.)

- Process birth/death causes implicit file open/close
- Oracle shadow processes open each datafile - but only at first access
 - Example. A Shadow Process that connects to perform a table scan for a report will open every datafile the tablespace is comprised of. When the query is completed and the process exits, each file will be implicitly closed through the internals of the exit() system call

Data
Access

Data
Protection

File Identification Feature (cont)

- ▼ With ODM, Oracle no longer uses file descriptors. Instead, ODM Identifiers are used.
- ▼ ODM Identifiers are shareable from process to process within the node
- ▼ Oracle caches ODM Identifiers in the SGA at instance startup
 - DBWR performs initial `odm_identify()` on all datafiles
 - CKPT identifies the control files and caches the information
 - LGWR is responsible for the initial identify and caching for REDO logfiles
- ▼ ODM Identifier usage reduces Kernel overhead
 - Calls to `exec()`, and `exit()` no longer cause implicit contention on Kernel file descriptor locks
 - Explicit contention is eliminated (`open()` / `close()`)

Data
Access

Data
Protection

Special Application - S.A.M.E

▼ Stripe And Mirror Everything

- Oracle's preferred DB physical layout method

▼ ODM is enabling technology for S.A.M.E.

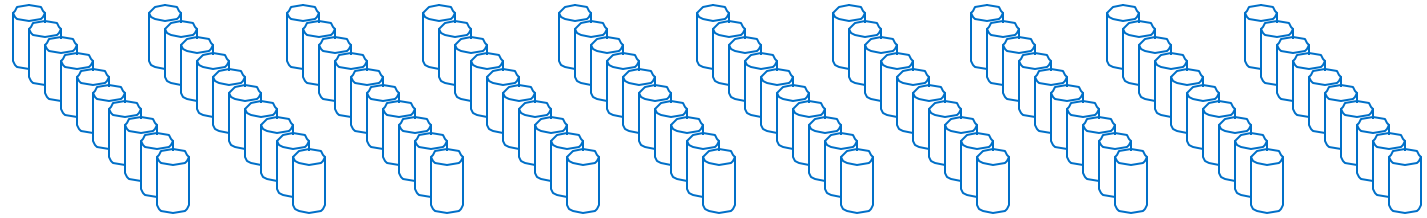
- In 9i, ODM files equal to raw partitions
- ODM is the best way to implement S.A.M.E methodology
- Large tables and indexes typically comprised of many partitions, therefore many datafiles.
- Deploying a database in accordance with S.A.M.E. methodology on raw partitions can be extremely cumbersome.
 - **Combining S.A.M.E methodology with large numbers of datafiles creates incredibly complex logical volume layouts**

Data
Access

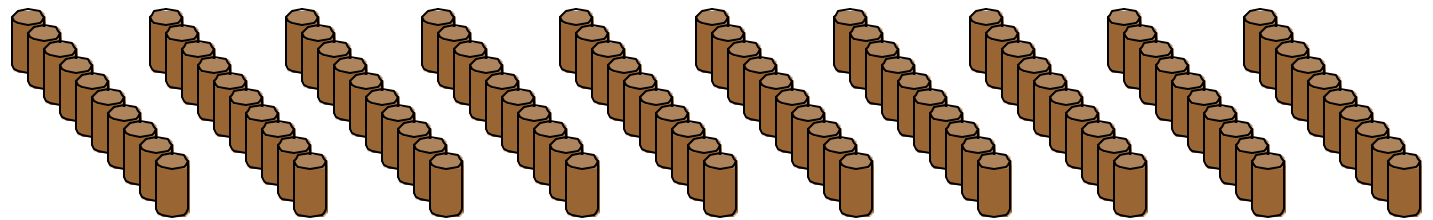
Data
Protection

Implementing SAME - Example

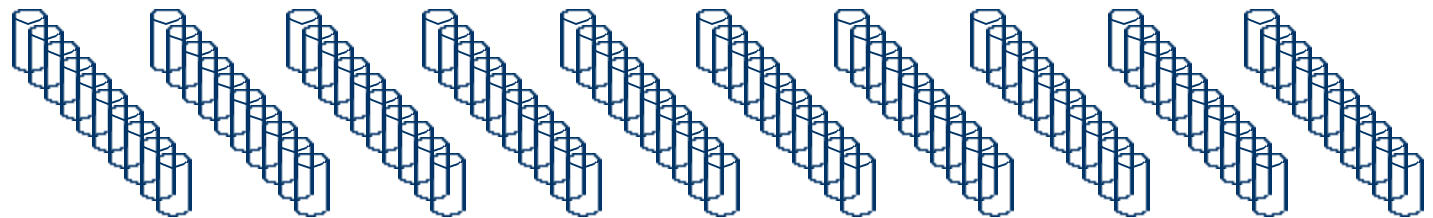
100 data files X



100 disks



means 100 raw partitions...



Data
Access

Data
Protection

Implementing SAME (contd.)

...then slice each disk into 100 subdisks for striping...

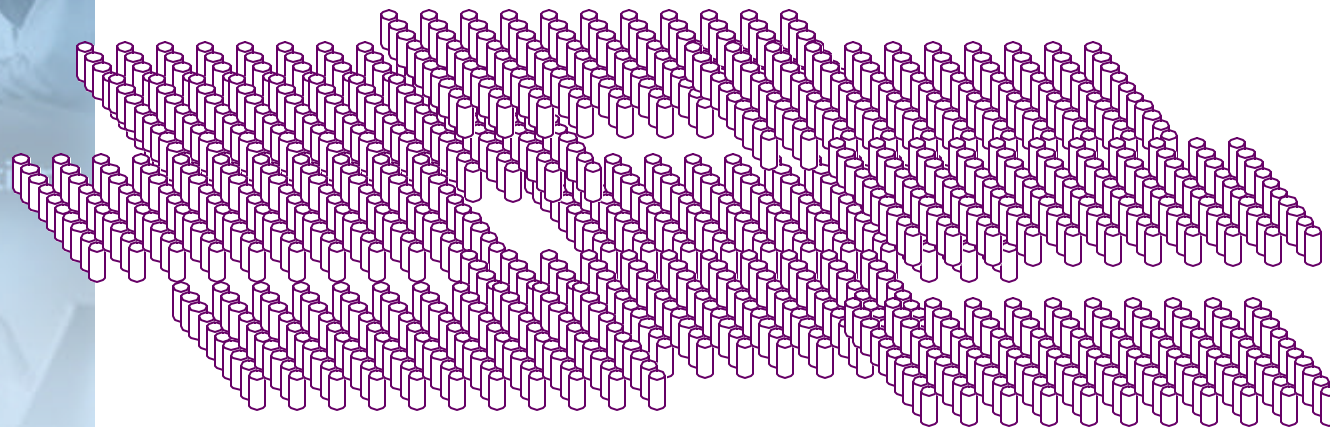
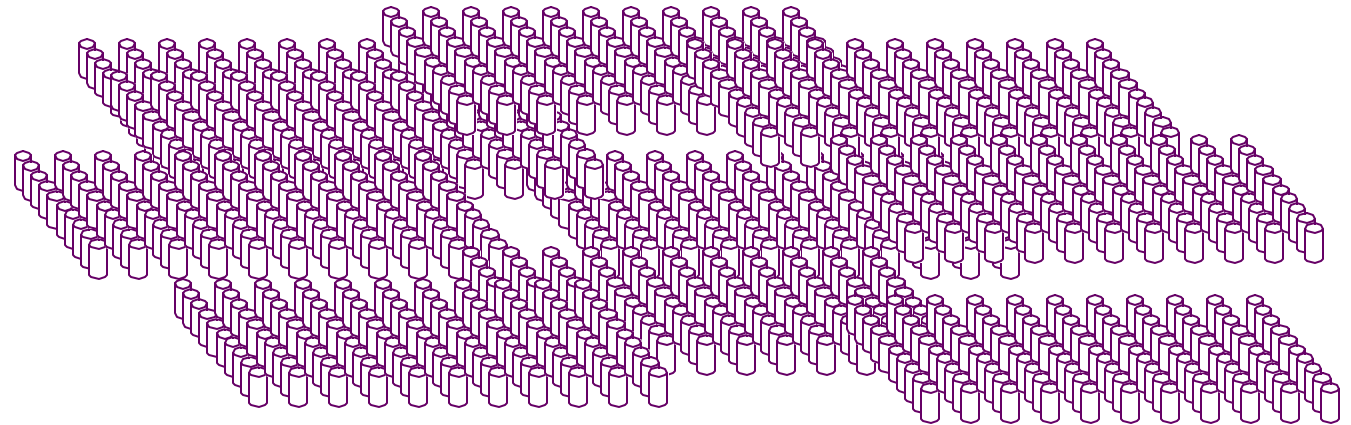


Data
Access

Data
Protection

Implementing SAME (contd.)

...and you get **ten thousand partitions.**
Then you have to start **mirroring.**

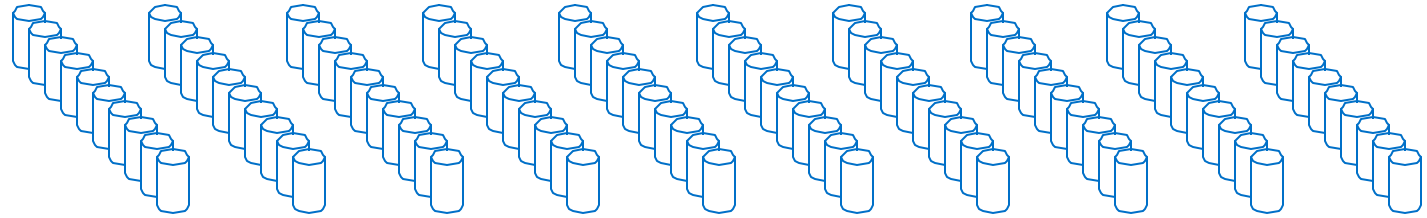


Data
Access

Data
Protect

Implementing SAME - Example

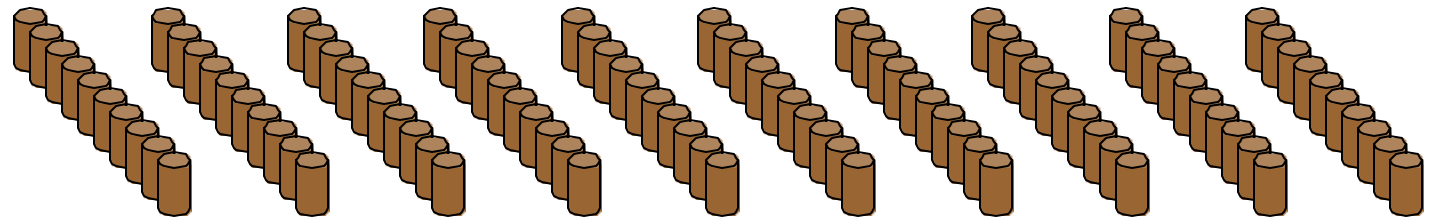
Under **ODM**, 100 data files...



...could reside in only **one** logical VxVM disk...



...striped across 100 physical disks. That's it.



Data
Access

Data
Protection

Quick I/O vs. ODM

Attribute	QIO	ODM
How exploited	Must be installed and set up	Automatically used if present (except for ufs)
How created contiguously?	<code>qiomkfile</code> utility	Done automatically by Oracle
Special backup requirements?	Yes. Must have QIO-friendly backup product	No. Appear just like regular files
Funny file name?	Yes, file name extension	No

Data Access

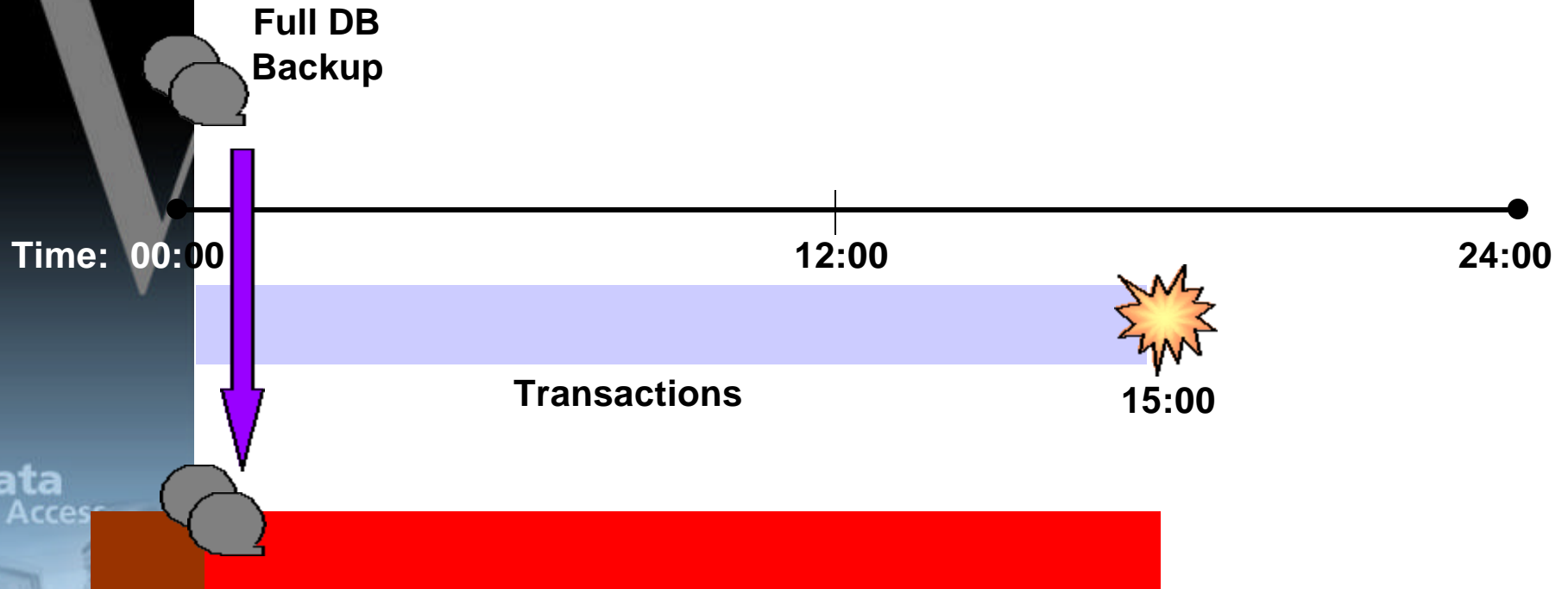
Data Protection

- ▼ **Less work**
- ▼ **Better availability**
- ▼ **Fewer headaches**
- ▼ **Less overhead**

Data
Access

Data
Protection

Traditional DB Recovery



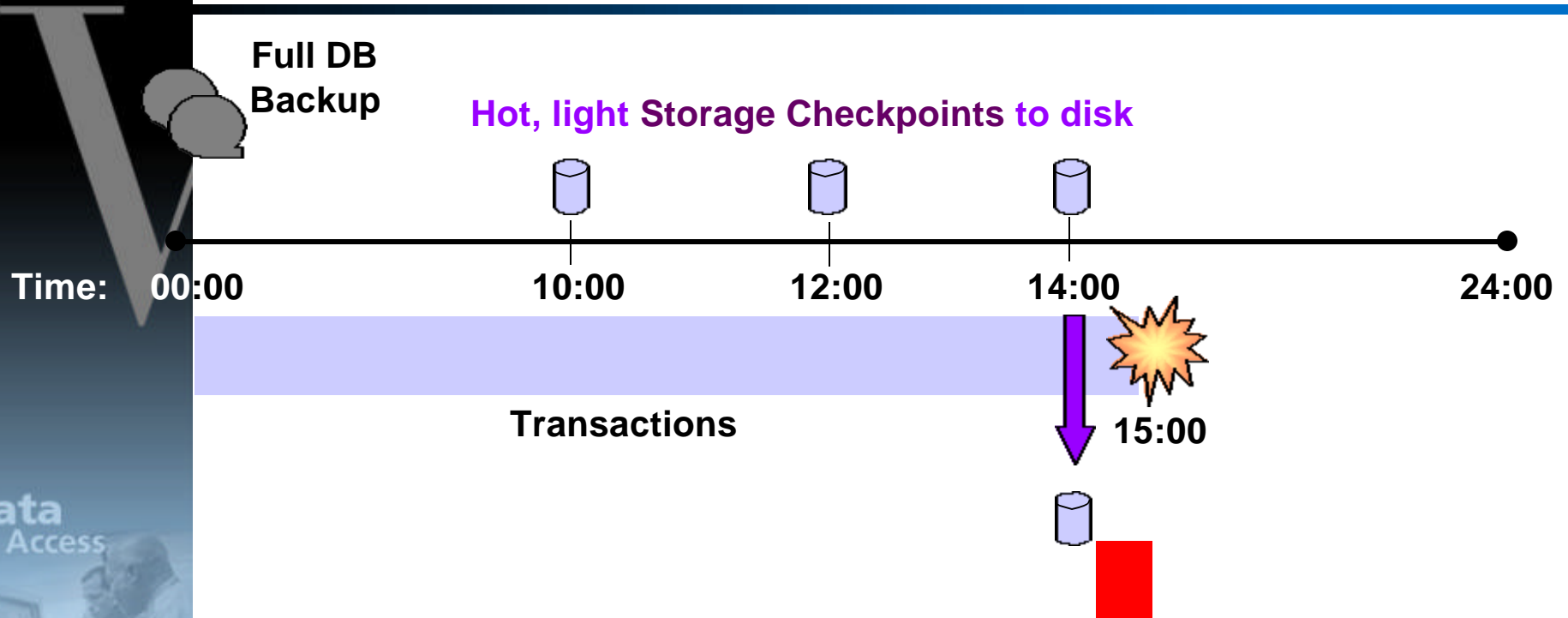
fsck processing (ufs)

Restore from backup

Reapply transactions 00:00 - 15:00

Back in service

Storage Checkpoint/Storage Rollback with VERITAS Database Edition *for Oracle*



Data
Access

Data
Protection

Restore from last checkpoint

Reapply transactions 14:00 - 15:00

Back in service - *fast*

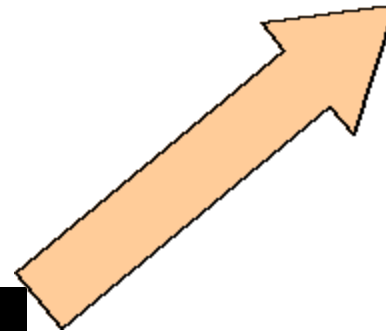
Improving Backup

Backup today must be:

- Invisible
- Fast
- Low-resource-consuming
- Unattended

**Hot
Light
Granular
Inexpensive**

**Cold
Heavy
Volume-Oriented
Expensive**



Data
Access

Data
Protect

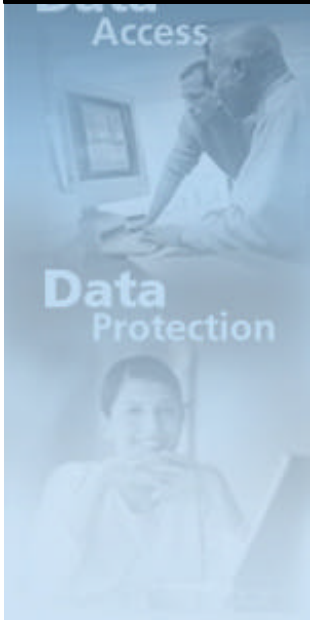
The “Perfect” Backup

Hot

Inexpensive

Light

Granular



Current Backup Offerings

Hot

VERITAS
Storage
Checkpoint/
Rollback

3rd
Mirror
Breakoff

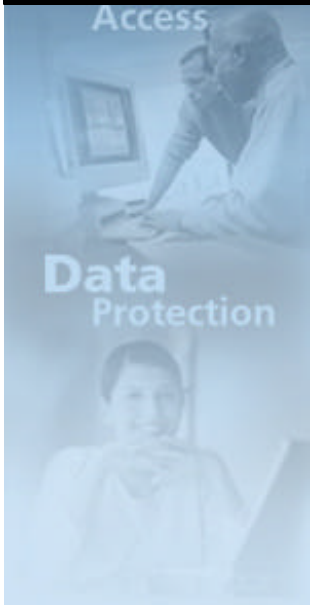
Inexpensive

Light

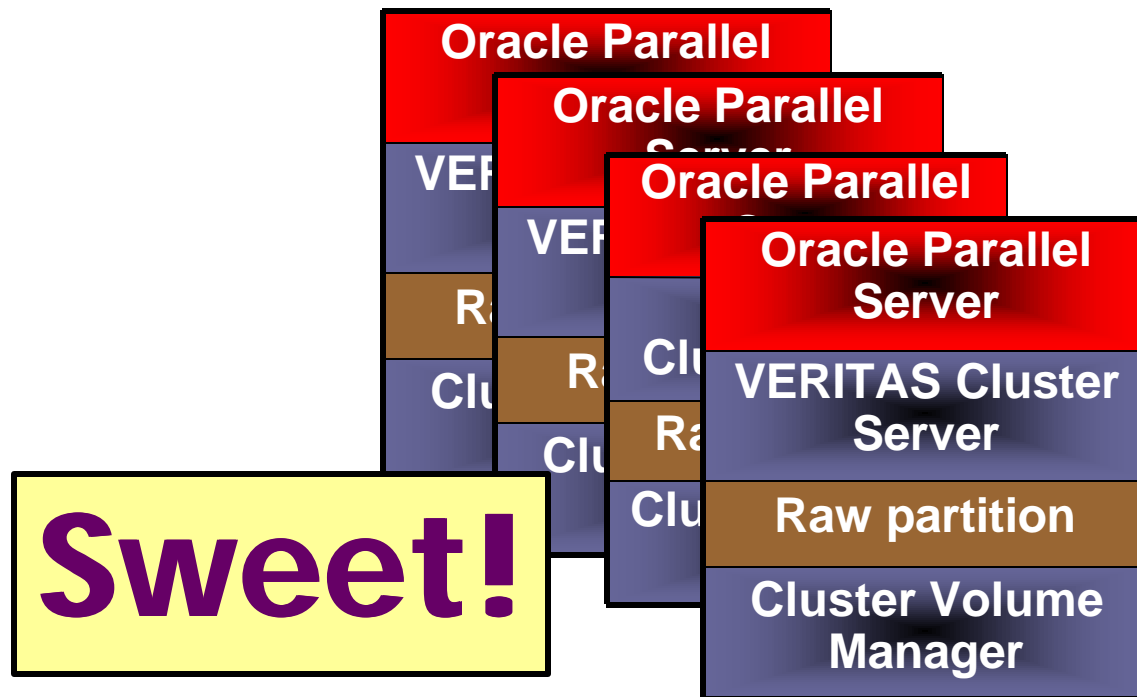
Dump

DB tools
(e.g. RMAN)

Granular



Database Edition/Advanced Cluster for Oracle



- ✓ Fast failover time for Oracle Parallel Server
- ✓ Support for a broad range of storage devices
- ✓ Beginning of the end of egregious vendor requirements

Data Access

Data Protection



Data
Access

Data
Protection

Election for
Oracle Datafile
Storage Type

Raw Partitions 3→

VxFS (With Raw Performance) 5→

ODM 7→

9→

11→

Other FS
Green Eggs
& Ham

8→

10→



Your Alternatives

- ❑ **I'm moving to 9i/ODM and Database Edition for Oracle 3.0 this year**
 - ❑ Your life is about to get better
- ❑ **I'm on Oracle 7/8, raw or ufs, and am not moving to 9i for the foreseeable future.**
 - ❑ Implement Quick I/O
- ❑ **I'm stuck in 32-bit still, with tons of memory I can't even address.**
 - ❑ Evaluate Cached Quick I/O
- ❑ **I'm on Oracle 3.6, and if it ain't broke don't fix it, I say.**
 - ❑ Man overboard!

Data
Access

Data
Protection

More Information

- ▼ See us here today
- ▼ See us on the web
<http://www.veritas.com/oracle>
- ▼ Oracle 9i information
 - <http://www.oracle.com>
 - Good luck finding ODM information ☹
- ▼ Oracle paper on S.A.M.E.
 - <http://technet.oracle.com/deploy/availability>
- ▼ Cached QIO performance paper
 - http://www.oracle.com/support/library/news/veritas_iss.html

Data
Access

Data
Protection



The End

Thank you for your time.

THE DATA AVAILABILITY COMPANY™